

## The Structure of the GroEL/GroES/ADP Complex

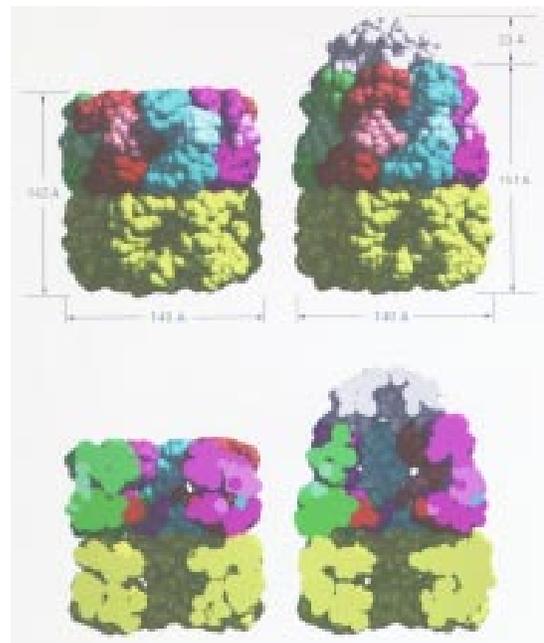
Z. Xu, A. Horwich and P. Sigler (Yale University)

The structure of the molecular chaperone GroEL in complex with a partner assembly GroES in the presence of bound ADP, was determined to a resolution of 3.0 Å using x-ray diffraction data collected at beamline X25 by Z. Xu, A. Horwich, and P. Sigler of Yale U. and the Howard Hughes Medical Institute (HHMI)<sup>[1]</sup>. The GroEL/GroES complex facilitates the folding of other proteins, and how this is accomplished is an important problem in modern biology. GroEL consists of two back-to-back 7-fold rotationally symmetric rings, that enclose two large, non-contiguous central cavities in which an unfolded protein can be bound. GroES consists of a single 7-fold rotationally symmetric ring, with mobile loops extending from its rim. In the presence of the nucleotide ATP, GroEL and GroES interact via these loops to form an asymmetrical GroEL/GroES complex. The x-ray

diffraction data show that the overall structure of this complex undergoes a substantial change upon the binding of the nucleotide, arising from movements of the domain components of one of the GroEL rings (**Figure B-1**). This creates an enlarged cavity from which the bound protein can be released, upon dissociation of the GroES ring from the GroEL. The diffraction data also provide insight to the binding of the nucleotide to the complex. The crystals of this complex have a very large unit cell and diffract x-rays very weakly and anisotropically, and moreover are relatively small. Access to the high beam intensity of X25 was mandatory in order to determine its structure to such high resolution. ■

[1] Z. Xu, A. Horwich, and P. Sigler, *Nature* **388**, 741-750, (1997).

**Figure B-1:** Overall architecture and dimensions of GroEL and GroEL-GroES-(ADP)<sub>7</sub>. Van der Waals space-filling models ( 6 Å spheres around C $\alpha$  ) of GroEL (**left**) and GroEL-GroES-(ADP)<sub>7</sub> (**right**). Upper panels are outside views, showing outer dimensions; lower panels show the insides of the assemblies and were generated by slicing off the front half with a vertical plane that contains the cylindrical axis. Various colors are used to distinguish the subunits of GroEL in the upper ring, with shading indicating domains; dark hue, equatorial domain; medium hue, apical domain; lightest hue, intermediate domain. The lower GroEL ring is uniformly yellow. GroES is uniformly gray.

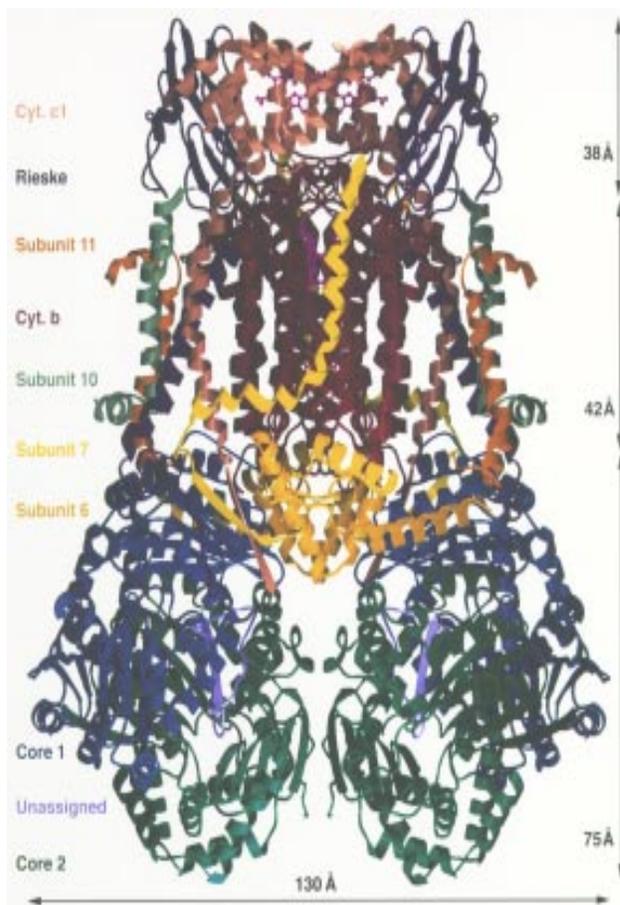


# Crystal Structure of Bovine Mitochondrial Cytochrome bc1 Complex, An Essential Component of Cellular Power Plant

D. Xia<sup>1</sup>, C. A. Yu<sup>2</sup>, H. Kim<sup>1</sup>, A. Kachurin<sup>2</sup>, L. Zhang<sup>2</sup>, L. Yu<sup>2</sup>,  
and J. Deisenhofer<sup>1</sup>

<sup>1</sup> HHMI & Department of Biochemistry, U. Texas, Southwestern  
Medical Center

<sup>2</sup> Department of Biochemistry, Oklahoma State U.



**Figure B-2:** Partial structural model of the dimeric cytochrome bc1 complex with polypeptides drawn as ribbons, hemes as stick models, and the Rieske iron-sulfur center as balls. Eight of the eleven subunits (core1, core2, cytochrome b, Rieske ISP, subunits 6, 7, 10, and 11) are completely traced and their sequences assigned. The top of the model is in the mitochondrial inter-membrane space, the middle spans the membrane, and the bottom is in the matrix space; approximate dimensions are given for each region.

All living cells use ATP (adenosine 5'-triphosphate) as a form of energy to sustain vital cellular processes essential for survival. ATP is constantly being generated in a cellular organelle called mitochondria — the cellular power plant, and the bc1 complex is one essential component of mitochondria directly involved in ATP synthesis. The bc1 complex performs an intermediate step in cellular respiration which involves a series of chemical reactions that frees energy in the form of ATP from food for use in cellular metabolism.

Cytochrome bc1 complex (ubiquinol-cytochrome c oxidoreductase, bc1) is the middle segment of the respiratory chain in almost all aerobic organisms, and an essential component in the photosynthetic machinery in purple bacteria. Green plants use highly homologous b6f complex as part of their energy photosynthetic apparatus. The bc1 complex is an integral membrane protein; it couples the electron transfer from ubiquinol to cytochrome c to the proton translocation across the membrane to generate a difference in pH and electrostatic potential to drive ATP synthesis. By determining the bc1 complex structure, we will be one step closer to a complete understanding of cellular of ATP generation.

Mitochondrial bc1 complex from bovine heart consists of 11 different subunits, two b- type hemes, one c-type heme and an iron-sulfur cluster with a total molecular weight of 250 Kd. The bc1 complex from bovine heart was purified

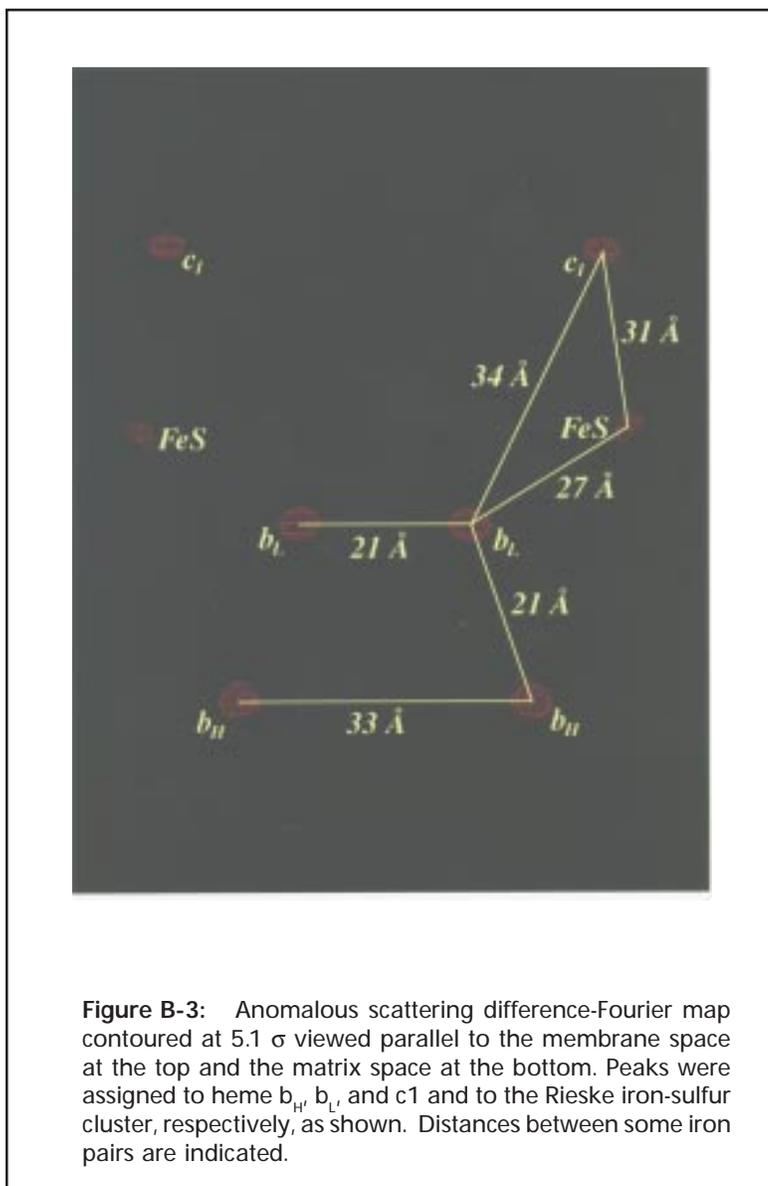
and crystallized, and bc1 complex crystals can be cryo frozen for stable data collection; they possess the symmetry of space group I4122 with cell dimensions of  $a=b=153.7$  Å and  $c=597.5$  Å. Synchrotron radiation sources was a critical factor in the successful structure determination. The use of beam time at beamline X12B of NSLS allowed data collection of both native and heavy atom derivatives in a relatively short period of time, a task that would have been impossible with a conventional x-ray source; the tunability of the x-ray wavelength at X4A enabled us to quickly and convincingly identify positions of iron atoms that are associated with redox-active prosthetic groups in the bc1 complex, and detect movement of the Rieske Iron Sulfur Protein (ISP); the high-flux x-ray beam at X25 permitted collection of a complete native data set to 2.7 Å resolution, making the refinement of bc1 structure possible.

The structure of bc1 was determined using MIRAS method with seven heavy metal derivatives. The current atomic model of the bc1 complex contains eight subunits completely sequence assigned, including core1, core2, cytochrome b, Rieske ISP, subunits 6, 7, 10 and 11; three subunits partially sequence assigned, including cytochrome c1, subunits 8 and 9; and four prosthetic groups which include two b-type hemes, one c-type heme and a 2Fe-2S cluster with a total number of amino acids residues in the model near 2000.

The bc1 complex forms a dimer in the crystal, and is 155 Å tall and 130 Å wide (**Figure B-2**). The whole complex can be divided into three regions: The membrane-spanning region is about 42 Å thick with 26 trans-membrane helices for the dimer; the cytochrome b dimer contributes 16 helices, cytochrome c1, the Rieske ISP, subunits 7, 10, and 11 account for the other five pairs of helices, respectively; projecting 38 Å into the inter-membrane space are the water soluble parts of the Rieske ISP and cytochrome c1, as well as subunit 8; the matrix region of the bc1 complex protrudes 75 Å into the mitochondrial matrix space and consists primarily of core1, core2 and subunits 6 and 9.

Taking advantage of the anomalous scattering property of iron atoms in the bc1 complex, we were able to determine the positions of and distances between these iron atoms in the bc1 complex (**Figure B-3**), and to postulate possible routes for electron transfer within the bc1 complex. More importantly, by combining anomalous diffraction experiments with specific inhibitor binding studies, we were able to detect very large conformational changes of the Rieske ISP upon binding of certain classes of inhibitors. These findings indicate a new mechanism for electron transfer from one subunit to another within the bc1 complex. ■

[1] D. Xia, C.A. Yu, H. Kim, J.-Z. Xia, A.M. Rachurin, L. Zhang, L. Yu, and J. Deisenhofer, *Science* 277, 60-66, (1997).



**Figure B-3:** Anomalous scattering difference-Fourier map contoured at  $5.1 \sigma$  viewed parallel to the membrane space at the top and the matrix space at the bottom. Peaks were assigned to heme  $b_H$ ,  $b_L$ , and  $c_1$  and to the Rieske iron-sulfur cluster, respectively, as shown. Distances between some iron pairs are indicated.

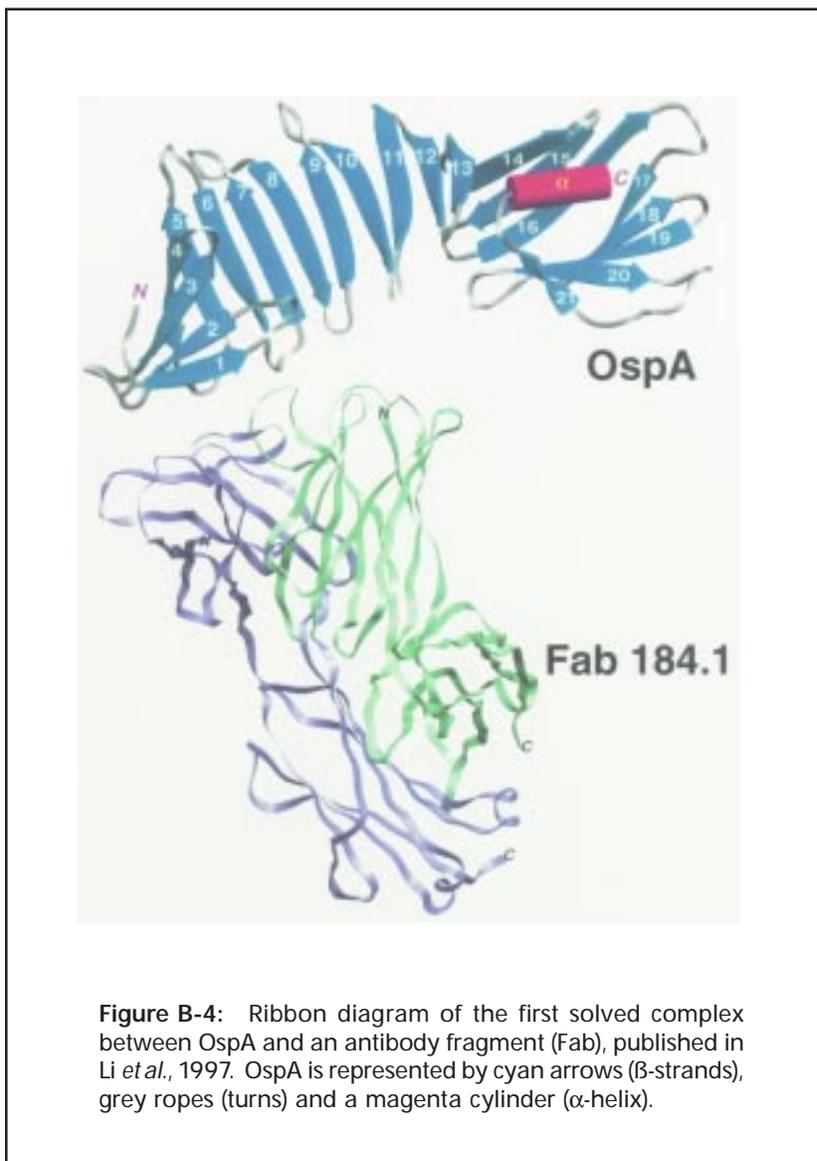
# Structural Lyme-ology

C. L. Lawson, H. Li, M. Becker, and W. Ding (BNL Biology Department)

The crystal structure of an antigenic protein from the outer surface of the Lyme disease bacterium has been determined by x-ray diffraction methods. OspA (outer surface protein A) is an abundant immunogenic lipoprotein of the Lyme disease spirochete *Borrelia burgdorferi* that is being developed as a vaccine to protect humans from being infected after a tick bite. Making use of data from beam lines X12C, X12B, and X25, the crystal structure of a recombinant form of OspA was solved in a complex with the Fab fragment of a mouse monoclonal antibody, and was refined to a resolution of 1.9 Å<sup>[1]</sup>. OspA is formed from a large antiparallel beta-sheet, with an unusual nonglobular region of “freestanding” sheet connecting globular N- and C-terminal domains (**Figure B-4**). Another unusual feature of the folding pattern in the nonglobular region is arrays of residues with alternating charges. A hydrophobic cavity buried in a positively charged cleft in the C-terminal domain is a potential binding site for an unknown ligand. The binding region of this antibody is biologically interesting because it overlaps with a well conserved surface in the N-terminal domain that is not accessible on the spirochete. This suggests that OspA interacts with other membrane proteins. The other end of OspA, the C-terminal domain, is what is exposed in the spirochete, and this, therefore, is the immunologically interesting part. Recently, native and tungstate-derivative MAD data have been collected at X12C and X25 on crystals of OspA complexed with the Fab fragment of an antibody that inactivates the Lyme spirochete. Preliminary interpretation of these data suggests that the antibody binds to a highly variable region of the C-terminal domain, which would have important implications for world wide effectiveness of OspA-based vaccines. Data also have been collected on the

highly homologous OspB lipoprotein, both alone and in a complex with the Fab fragment from an antibody that causes Lyme spirochetes to self-destruct. Multi-wavelength anomalous diffraction (MAD) data will soon be collected on tungstate-derivative OspB crystals. Patterson maps calculated from a single wavelength dataset at the tungstate L-III edge have already pinpointed the heavy-atom binding sites. ■

[1] H. Li, J.J. Dunn, B.J. Luft, and C.L. Lawson, *Proc.Nat. Acad. Sci.USA* **94**, 3584-3589, (1997).



# The Structure of the Human Natural Killer Cell Inhibitory Receptor

Q. Fan, L. Mosyak, and D.C. Wiley (Harvard University)

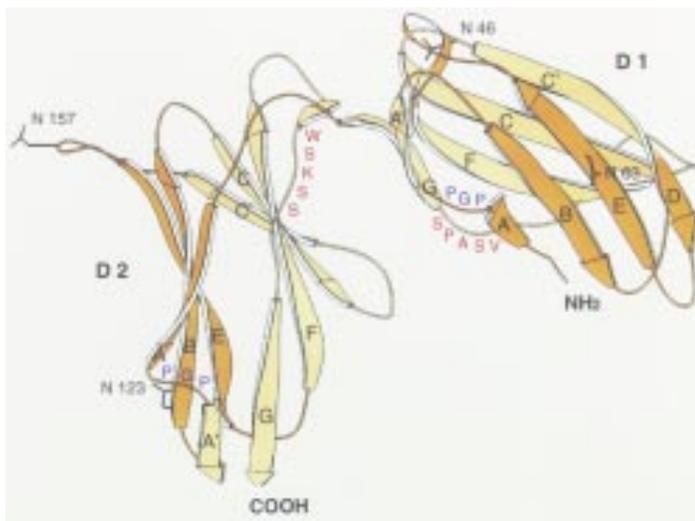
Natural killer (NK) cells lyse abnormal cells deficient in class I major histocompatibility complex (MHC) expression. Killer cell inhibitory receptors (KIR) on human NK cells prevent the lysis of target cells bearing specific class I MHC molecules. The x-ray structure of a p58 KIR (p58-cl42) was determined at 1.7 Å resolution using the multiwavelength anomalous diffraction (MAD) method. MAD data of a single selenomethionyl p58-cl42 crystal were collected to 2.2 Å with a 300 mm diameter MAR Research image-plate system on the X25 beamline of the NSLS. Based on the x-ray absorption spectrum of the Se-Met p58-cl42 crystal, three wavelengths for data collection were chosen, which included one wavelength at the absorption peak (0.9791 Å), one at the inflection point (0.9794 Å), and a third at a remote high energy point (0.9639 Å). A native data set was collected to 1.7 Å on F-1 beamline at CHESS. MAD phasing was treated as a case of multiple isomorphous replacement. The initial MAD map was improved by

density modification. A model of p58-cl42 KIR was traced from both the density modified and unmodified electron density map. The atomic model has been refined against the native data set at 1.7 Å resolution to a crystallographic R-value of 20.6% ( $R_{\text{free}} = 25.4\%$ ).

The p58-cl42 KIR structure has tandem immunoglobulin-like domains positioned at an acute, 60 degree angle (**Figure B-5**). Loops on the outside of the elbow between the domains form a binding site projected away from the NK cell surface. The topology of the domains and their arrangement relative to each other reveals a relationship to the hematopoietic receptor family. We propose, by analogy to the hematopoietic receptors, that signalling by KIR may involve receptor dimerization. ■

[1] Q.R. Fan, L. Mosyak, C.C. Winter, N. Wagtmann, E.O. Long, and D.C. Wiley, *Nature* **389**, 96-100, (1997).

**Figure B-5:** Structure and topology of p58-42 KIR. KIR domain D1 is N-terminal, and D2 is C-terminal; ABED β-sheet (dark orange) and C'CFG A' β-sheet (light orange); WSKSS and VSAPS motifs (red); potential glycosylation sites at asparagines 46, 63, 123, and 157 (black).



# The Structure of the Cre Recombinase/DNA Complex

F. Guo, D.N. Gopaul, and G.D. Van Duyne (University of Pennsylvania)

Genetic recombination involves the breaking and rejoining of DNA strands between two sites to provide a new strand connectivity and therefore an alternative organization of the genetic code. Organisms use DNA recombination for a number of purposes, including the generation of genetic diversity, the repair of damaged DNA, the integration and excision of viral genomes into and out of the host chromosome, the regulation of gene expression, and in ensuring the stable inheritance of circular chromosomes and plasmids. The lambda integrase family of enzymes that mediates recombination in bacteria and yeast recognizes specific DNA sequences and carries out the exchange of DNA strands only at these sites<sup>[1]</sup>. The bacteriophage P1 Cre recombinase, a member of the lambda integrase family, carries out site-specific recombination between 34-base pair DNA sequences called loxP sites<sup>[2]</sup>. The simplicity of the Cre/loxP system (no other proteins or DNA sequences are required) has led to a rapid rise in its use as a tool in a number of

genetic engineering experiments, particularly those involving the study of gene function in transgenic mice<sup>[3]</sup>.

The work described here involves the recombination reaction between Cre recombinase and its DNA substrate, the loxP site. A cartoon representation of this reaction is shown in **Figure B-6**. Two Cre molecules bind to each of the two loxP substrates and bring the double-stranded helices together to form a recombination synapse. Cre then cleaves two of the four DNA strands in the synapse, using a tyrosine side chain as a nucleophile. Transient phosphotyrosine links between protein and DNA are formed, thereby conserving the energy of the phosphodiester bond. The free DNA strand produced upon cleavage then acts as a nucleophile, attacking the phosphotyrosine linkage located across the synapse on the partner substrate. This results in the exchange of two of the four DNA strands and formation of a Holliday junction intermediate. A complementary set of cleavage and strand exchange steps on the remaining pair of DNA

strands then completes the recombination event and gives recombinant DNA duplexes.

In work carried out at the X25 and X4A beamlines at the NSLS, G. Guo, M. Gopaul, and G. Van Duyne from the University of Pennsylvania School of Medicine have trapped and determined the three-dimensional structure of a Cre-loxP reaction intermediate where the recombinase has cleaved the loxP substrate to form a phosphotyrosine linkage, but is unable to complete the strand exchange step<sup>[4]</sup>. To accomplish this, they formed and crystallized a 200 kD synaptic Cre-DNA complex using a suicide loxP substrate that loses its attacking nucleotide upon cleavage. Crystals of this Cre-DNA complex diffracted X-rays poorly on home sources and required the wiggler source and optics at X25 in order to obtain the relatively high resolution (2.4 Å) for a structure of this complexity. In addition, the resources of the X4A beamline were used to conduct a multiwavelength

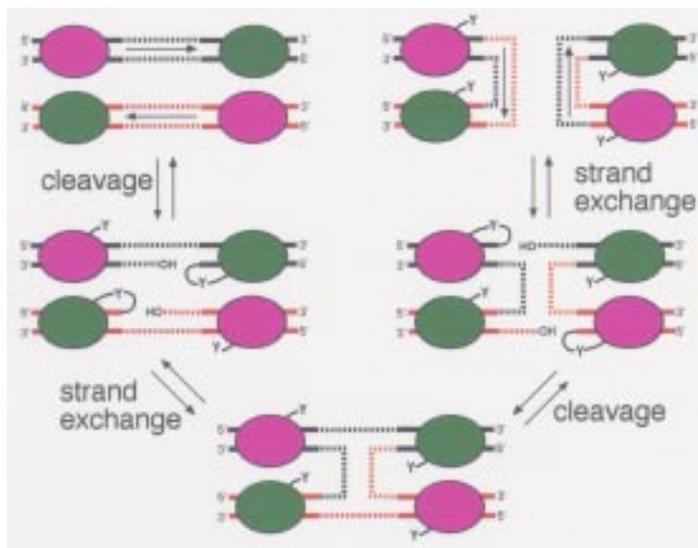


Figure B-6: Representation of the cleavage, exchange, and recombination processes involved in the breaking and rejoining of two DNA double helices, mediated by the bacteriophage P1 Cre recombinase. Two successive cycles of strand cleavage, exchange, and recombination are required to fully carry out the recombination event.

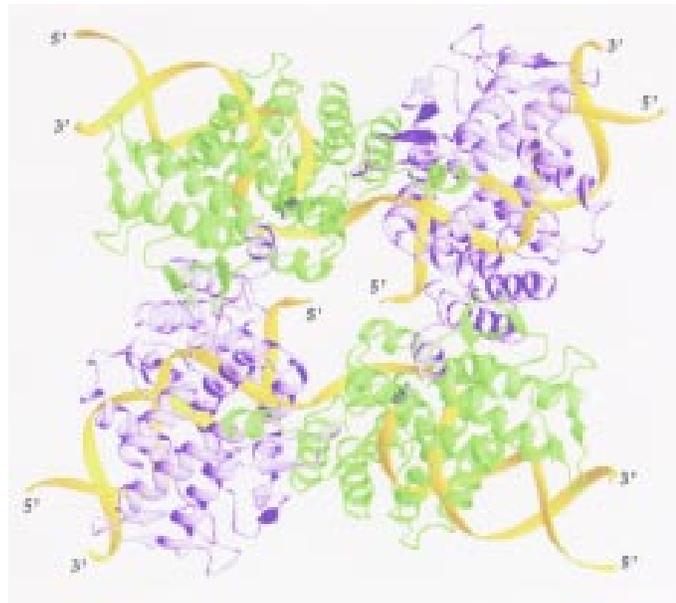
anomalous diffraction (MAD) experiment using selenomethionine-containing Cre recombinase at the Se absorption edge. The anomalous scattering contribution from 22 selenium atoms in the crystallographic asymmetric unit was crucial in obtaining crystallographic phases to interpret the structure.

The Cre-DNA intermediate structure is shown in **Figure B-7**. Four recombinase molecules (green and purple ribbons) form a pseudo-fourfold symmetric tetramer that is held together by a network of protein-protein interactions. The cleaved loxP sites (gold ribbons) have been bent by about 90° and adopt a nearly square planar configuration. The green recombinase molecules have cleaved the loxP site and formed phosphotyrosine bonds to the DNA. The purple recombinase molecules have not cleaved the DNA substrates and one of the interesting questions addressed by this structure involves understanding how the recombinase tetramer coordinates the cleavage reactions. The free ends of the cleaved DNA strands have been pushed towards the center of the synaptic complex, in what appears to be the start of the strand transfer process. The Cre-DNA synaptic intermediate structure has provided the first three-dimensional framework for understanding many aspects of the lambda integrase family site-specific recombination

reaction and has both supported a number of recent proposals about how the reaction occurs and has provided a fresh view of other aspects. The structural model also promises to serve as an important starting point for the design of improved recombinases and substrates for use in genetic engineering applications. ■

- [1] N.L. Craig, "The Mechanism of Conservative Site-Specific Recombination", *Annual Review of Genetics* **22**, 77-105, (1988).
- [2] N. Sternberg, D. Hamilton, S. Austin, M. Yarmolinsky, & R. Hoess, "Site-Specific Recombination and its Role in the Life Cycle of Bacteriophage P1", *Cold Spring Harbor Symposia on Quantitative Biology* **1**, 297-309, (1981).
- [3] B. Sauer, "Manipulation of Transgenes by Site-Specific Recombination: use of Cre Recombinase", *Methods in Enzymology* **225**, 890-900, (1993).
- [4] F. Guo, D. N. Gopaul, G. D. Van Duyne, "Structure of Cre Recombinase Complexed with DNA in a Site-Specific Recombination Synapse", *Nature* **389**, 40-46, (1997).

Figure B-7: Ribbon representation, based on x-ray diffraction results, of the Cre-DNA intermediate formed following DNA cleavage, but for which the strand exchange process was not completed. DNA strands are shown as gold ribbons, and the Cre recombinase molecules are shown as green and purple ribbons. The cleaved DNA loxP sites have been bent to form a nearly square planar configuration, and the free ends of the cleaved strands point toward the center of the complex, to strand exchange.



# Crystal Structure of the P4-P6 Domain from a Group Ribozyme Reveals Principles of RNA Packing

J. Cate and J. Doudna (Yale University)

The discovery that RNA can function as a biological catalyst, or ribozyme, has raised important questions about the chemistry and evolution of enzymes. Several classes of ribozymes have been identified and characterized, but until recently there was little information about the structure of RNA. The sequences of most families of ribozymes consist of a conserved series of short base-paired stems (helices) connected by “loop” regions. Biochemical experiments suggested that the helices of large RNA molecules often pack into globular structures with solvent-inaccessible cores, much like proteins.

Group I self-splicing introns, the first class of ribozymes discovered, have the remarkable ability to excise themselves from precursor transcripts without the aid of any protein. The structure of the RNA holds the key to understanding the self-splicing mechanism as well as evolutionary relationships among ribozymes. The group I intron from *Tetrahymena thermophila* has been well characterized biochemically, revealing the ribozyme to be comprised of two structural domains. The domains can be synthesized separately and assembled in trans to produce active ribozymes. The 160-nucleotide P4-P6 domain, containing half of the residues of the catalytic core, is an independently folding structural unit. Following the development of techniques for RNA crystallization, crystals of the P4-P6 domain were obtained that diffracted X-rays to better than 2.8 Å resolution.

The solution of the P4-P6 crystal structure was achieved through an insightful observation by Jamie Cate, then a graduate student working on the project at Yale. Jamie noticed that P4-P6 crystals grew best in the presence of tiny amounts of cobalt hexammine, a small molecule that mimics the geometry of a fully hydrated magnesium ion. Since dramatic effects on crystal growth were seen with stoichiometric amounts of hexammine to RNA, Jamie reasoned that there might be just a few specific hexammine binding sites present in the RNA. To take advantage of such potential sites for making heavy atom derivatives, Jamie obtained from Prof. Henry Taube at Stanford a sample of osmium hexammine, and he soaked this compound into the P4-P6 crystals. An anomalous difference Patterson map calculated using diffraction intensities measured on a laboratory x-ray source showed

several strong peaks whose positions could be refined and cross-checked. The osmium derivative was used to solve the P4-P6 crystal structure by multiwavelength anomalous diffraction (MAD) with data measured at NSLS beamline X4A (**Figure B-8**).



**Figure B-8:** 2.8 Å crystal structure of the 160-nucleotide P4-P6 of the *Tetrahymena* self-splicing ribozyme. Helices of the conserved catalytic core, in blue and red, pack against helices of an extended region. Tertiary contacts occur between an A-rich corkscrew motif (orange) and the P4 helix (blue), and between a tetraloop (gold) and its receptor (green).

The 2.8 Å crystal structure of the P4-P6 domain is the first crystal structure of an RNA molecule large enough to show side-by-side packing of helices. Such packing is thought to be necessary to form stable active sites as found in self-splicing introns, Ribonuclease P, the ribosome and the spliceosome. In the domain, a sharp bend allows stacked helices of the conserved core to pack against helices of an adjacent region. Two specific long-range interactions clamp the two halves of the domain together: a magnesium-coordinated adenosine-rich corkscrew plugs into the minor groove of a helix, and a GAAA hairpin loop binds to a conserved 11-nucleotide internal loop. Metal- and ribose-mediated backbone contacts further stabilize the close helical packing, giving rise to a solvent-inaccessible interior.

One interesting feature of the P4-P6 structure is the discovery of an unexpected new motif that mediates both intra- and intermolecular interactions. At three separate locations in the P4-P6 molecule, adjacent adenosines in the sequence lie side-by-side, stacking simultaneously on both sides of the helix. This creates an adenosine platform, opening the minor groove for stacking or base-pairing with nucleotides from a distal RNA strand. This motif explains in part the preponderance of adenosines in internal loops of many RNAs.

Most of the contacts that stabilize internal P4-P6 domain structure as well as the packing of arrays of molecules in the crystal lattice involve the minor groove. The wide and shallow minor grooves of A-form helices are generally more accessible than the deep major grooves.

However, non-canonical base pairings, loops and bulges in RNA can perturb the geometry of a helix, affecting the shape and electrostatic potential in localized regions. Indeed, local perturbations at tandem G-U “wobble” base pairs in the P4-P6 RNA give rise to specific metal binding pockets in the major groove: the hexamine binding sites used in the structure determination. In two cases these sites are occupied by fully-hydrated magnesium ions in the native structure. The binding sites expand the known repertoire of ligand binding motifs in RNA and may provide a general means of derivatizing appropriately designed RNAs for x-ray crystallographic analysis. Phylogenetic sequence comparisons show that one class of these motifs occurs frequently in ribosomal RNAs, suggesting a possible mechanism for metal binding in the ribosome. ■

[1] J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, and J.A. Doudna, “Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing”, *Science* **273**, 1678-85, (1996).

[2] J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, A. Szewczak, C.E. Kundrot, T.R. Cech, and J.A. Doudna, “RNA Tertiary Structure Mediation by Adenosine Platforms”, *Science* **273**, 1696-9, (1996).

[3] J.H. Cate, and J.A. Doudna, “Metal Binding Sites in the Major Groove of a large Ribozyme Domain”, *Structure* **4**, 1221-9, (1996).

# The Structure of a Component of the AIDS Virus: the Carboxyl-Terminal Dimerization Domain of the HIV-1 Capsid Protein

T. R. Gamble, S. Yoo, F.F. Vajdos, U.K. von Schwedler, D.K. Worthylake, H. Wang, .P. McCutcheon, W.I. Sundquist, C.P. Hill (University of Utah)

As reported recently<sup>[1]</sup> we have solved the x-ray crystal structure of the C-terminal dimerization domain of the capsid protein from the human immunodeficiency virus type 1 (HIV-1). The 26-kD capsid protein (CA) of HIV-1 results from the action of the well-known HIV protease: as the virus buds, a 55-kD polyprotein called Gag is processed by the viral protease to produce three discrete new proteins — one, called matrix, that binds to the viral membrane, the capsid protein which forms the conical viral core structure, and a third called nucleocapsid that helps to organize the viral RNA — as well as several smaller peptides. These are the major structural proteins of HIV, with ~2, 000 copies of each protein present in a single viral particle.

CA performs several essential roles during the HIV life-cycle. Most importantly, CA forms the conical core structure that surrounds the viral genome at the center of the mature virus. Genetic analyses indicate that Capsid also functions during assembly of new virus particles and

as the virus enters a new host cell. Our study focused on the COOH-terminal domain of Capsid, residues 146-231 (CA146-231). This domain contains a stretch of 20 amino acids, termed the MHR (major homology region), that is found in all known onco- and lentiviruses, and in the yeast retrotransposon Ty-3. As one might expect, this highly conserved sequence is essential for viral replication, although its precise functions remain unclear. Another feature of special interest is the role of CA146-231 in mediating Capsid dimerization; this domain dimerizes with the same affinity as the full length protein. Capsid dimerization is essential for core formation and infectivity, and the assembly of infectious viral particles.

In an effort to understand the structural basis for Capsid dimerization and MHR conservation, we determined the crystal structures of recombinant selenomethionine-substituted CA151-231 at 1.8 Å resolution and native (no Se) CA146-231 at 3.1 Å resolution. Diffraction data for the CA151-231

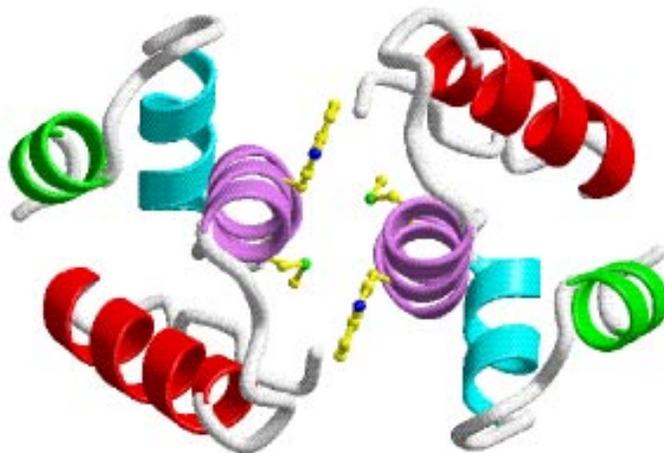


Figure B-9: Ribbon diagram of the structure of the HIV-1 capsid domain CA(146-231) viewed as a dimer. Two residues required for stabilization of the dimer *in vivo* are shown as stick models near the center of the figure.

crystal-structure analysis were taken at NSLS beamline X12C. The structure was solved by the multiwavelength anomalous diffraction method (MAD), based on scattering from Se atoms that had been introduced to the protein in methionine residues. The structure determination was relatively rapid, with an electron-density map produced and an essentially complete model built only 54 hours after data-collection began. Residues 151-219 of CA151-231 formed a well ordered globular domain in this structure. The standard crystallographic R-value is 22.9% and the free R-value (used for cross validation of the result) is 27.4% at the end of the refinement of the structure of CA151-231. Once the CA151-231 structure was known, it was used to determine the CA146-231 structure by the method of molecular replacement.

The CA146-231 and CA151-231 structures are very similar, with each molecule composed of an extended strand followed by four  $\alpha$ -helices. **Figure B-9** shows the structure of the biologically relevant CA146-231 dimer that is formed in the crystals. In this figure, the MHR is composed of the helix shown at the lower left (or upper right) and the strand that lies behind it in the figure. The role of the most highly conserved MHR residues is to form a hydrogen-bonding network that stabilizes this structure and links it to the helix that is seen end-on near the center of the dimer. The structure therefore explains the observation that certain "conservative" mutations in the MHR, for example Gln to Asn or Glu to Asp, block HIV-1 replication; even mutational changes as small as the deletion of a single carbon atom appear to disrupt the hydrogen-bonding network within the MHR.

A series of four conserved hydrophobic residues also are essential for MHR function; they lie on one surface of this helix and contribute to the hydrophobic core of the protein. Thus, all of the conserved MHR residues perform critical structural roles. Although most of the Gag polyprotein is quite variable among other viruses, the remarkable conservation of this 20-residue segment suggests that the MHR structure mediates an essential interaction with a highly conserved binding partner, such as a cellular factor or an invariant segment of a viral protein.

CA146-231 possesses almost identical dimerization affinity as the full length CA protein. The dimer interface is created by the parallel packing of two helices, to create a hydrophobic core. It is notable that the dimer interface and MHR are distinct non-overlapping structures. Site-directed mutagenesis was used to confirm that the dimer seen in the CA146-231 crystal is also the dimer interface of full-length capsid protein in solution. Two interface residues, Trp184 and Met185 (see **Figure B-9**), were mutated to Ala in the intact CA. In neither case would the mutant protein form dimers in solution, thereby indicating that the crystallographic structure is indeed the authentic dimer. One of these mutations allowed the formation of viral particles in cell cultures, although the particles produced were not infectious. This indicates that the crystallographic dimer is an essential interaction in the assembly and maturation of viral particles, and provides another possible target for the design of drugs to treat AIDS. ■

[1] T.R. Gamble *et al.*, *Science* **278**, 849 - 853 (1997).